REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED
08-30-2002	Final Report	01-01-1999 to 12-31-1999

4. TITLE AND SUBTITLE
Rapid PCR Assays that Specifically Identify Anthrax and Anthrax Surrogate Chromosomal Signatures

5a. CONTRACT NUMBER

5b. GRANT NUMBER
N00014-99-1-0303

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Leighton, Terrance

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Department of Biochemistry and Molecular Biology
401 Barker Hall
University of California
Berkeley, California 94720-3202

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217-5000

10. SPONSOR/MONITOR'S ACRONYM(S)
ONR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AILABILITY STATEMENT
Distribution Unlimited

20021104 043

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The genetic variation among a set of 175 full-length sspE DNA sequences obtained from representative members of the B. anthracis clade have been examined. Thirty-six sspE genotypes and seventeen protein phylotypes were identified among the B. cereus, B. thuringiensis, B. anthracis and B. mycoides strains analyzed. The most extensive genetic diversity was seen in B. thuringiensis strains representing 78 serovar classes. Computational analysis of the sspE DNA sequence data sets suggests that the B. anthracis clade is more phylogenetically complex than has been inferred by traditional taxonomic methods.

15. SUBJECT TERMS
Anthrax, PCR assays, genetic diversity, molecular phylogeny

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Terrance Leighton
Unclass.	Unclass.	Unclass.	UL	6	19b. TELEPHONE NUMBER (Include area code) 510 642-1620

FINAL REPORT

Grant Number: N00014-99-1-0303

Principal Investigator: Terrance Leighton

Institution: University of California Berkeley

Grant Title: Rapid PCR Assays that Specifically Identify Anthrax and Anthrax Surrogate Chromosomal Signatures

Award Period: 1/1/99 - 12/31/99

OBJECTIVE: The goal of this project is to develop rapid DNA-based assay systems targeted to anthrax and anthrax surrogate spore structural signatures. To mitigate assay false alarms, these detection systems will be designed based on a detailed understanding of the phylogenetic landscape containing the target organism and its near-neighbor species. These systems will enable the specific identification of vaccine, non-pathogenic, genetically engineered and fully virulent anthrax strains. Parallel technologies will be developed for the anthrax surrogate, *Bacillus thuringiensis*, to allow the rigorous evaluation of DNA-based detection technologies under field conditions.

APPROACH: This proposal focuses on developing rapid PCR assays for *B. anthracis* and its near-neighbors by sequencing full-length *sspE* spore core genes and DNA regions flanking the *sspE* coding sequence. This DNA sequence dataset will facilitate the development of anthrax assays with improved specificity. We will also use these data to initiate development of rapid PCR assays that can specifically identify *B. cereus*, *B. mycoides* and *B. thuringiensis*. To further document the specificity of the *sspE* assay, we will sequence the *B. cereus*, *B. thuringiensis* and *B. mycoides* genomic *sspE* region from geographically diverse culture collections maintained by the ATCC, Bacillus Genetics Stock Center, Institut Pasteur and DSM.

ACCOMPLISHMENTS: We have obtained 175 full-length *sspE* DNA sequences from representative members of the *B. anthracis* clade. Computational analyses of the *sspE* DNA sequence data sets suggest that the *B. anthracis* clade is more phylogenetically complex than had been inferred by traditional classification methods. Thirty-six *sspE* genotypes and seventeen protein phylotypes were identified among the *B. cereus*, *B. thuringiensis*, *B. anthracis*, and *B. mycoides* strains analyzed. The most extensive genetic diversity was seen in *B. thuringiensis* strains representing 78 serovar classes. In many cases, strains within a *B. thuringiensis* serovar showed identical SspE sequence suggesting a significant correlation of SspE sequence clustering with serotyping.

The *sspE* amplicons obtained contain ribosome binding site sequences,

translational initiation codons, complete protein coding sequences, and translational termination codons. All of the *B. anthracis* group *sspE* genes have a near canonical Shine-Dalgarno ribosome binding site sequence, GGAGGT, followed by an ATG translation start codon. Coding sequences are translated into SspE proteins of 92-95 amino acids. All of the *sspE* genes terminate translation with at least two tandem stop codons, TAATGA. A *B. thuringiensis sspE* gene that terminates translation with three contiguous stop codons, TAGTAATGA, has been also identified. The use of an ATG start codon and multiple contiguous stop codons are regulatory motifs typical of highly expressed genes supporting the finding that SspE is a highly abundant sporulation protein in *Bacillus* cells.

A universal PCR primer combination was developed that was able to amplify approximately 279 base pairs within the *sspE* coding regions of all experimental strains studied. The sequence of this PCR amplicon was different in each of the classically defined *Bacillus* species that were examined. Translated amino acid sequences of these fragments contained the majority of amino acid alterations included among SspE proteotypes. This universal primer combination provides a simple method for the unambiguous discrimination of ecological populations within the *B. anthracis* group.

CONCLUSIONS: We have demonstrated the utility of SspE as a molecular chronometer for the phylogenetic positioning of ecologically distinct populations within the closely related *B. anthracis* taxa. The genetic variation among a set of 175 *B. anthracis* clade full-length *sspE* DNA sequences has been analyzed. In many cases, the *sspE* phylotypes naturally clustered with pathotypes, serotypes and ecotypes identified by independent methods. These results support the hypothesis that sequences of informative protein-coding genes are more effective in distinguishing ecologically distinct populations than either DNA-DNA hybridization or rDNA gene sequences.

SspE phylogeny suggests that *B. cereus* and *B. thuringiensis* are more closely related to each other than to *B. anthracis* and *B. mycoides*. These findings are consistent with the AFLP-based molecular variation studies carried out previously by Keim *et al.* In addition, a pathotype analysis of *B. cereus* group virulence factor genetic variation by Guttmann and Ellar also suggested that *B. cereus* and *B. thuringiensis* are more closely related to each other than they are to *B. anthracis*.

SIGNIFICANCE: The results presented here establish that the *sspE* gene set possesses emergent DNA sequence properties which enable the systematic study of natural variation within the *B. anthracis* group. A particularly powerful aspect of SspE phylogenetic reconstruction is the ability to generate self-ordering systematic data properties that naturally position *B. anthracis* taxa into stratifying groups. These attributes allow the creation of an internally consistent interpretative framework for the phylogenetic analysis of *B. anthracis* group microorganisms. These studies also extend Cohan's previous demonstration of the utility of other protein-coding genes for the phylogenetic separation of closely related groups within the *Bacillus* genus.

PATENT INFORMATION: No patents have been filed

AWARD INFORMATION: Dr. El Helow was awarded a Fulbright Faculty fellowship

PUBLICATIONS and ABSTRACTS (for total period of grant):

1. Leighton, T., Helow, E. E., Fouet, A. and M. Mock. Molecular Phylogenetic and Ecological Analysis of the *B. anthracis* Clade. 1999. DP99, 1st European Conference on Dangerous Pathogens, Winchester, England.

2. Eisenstadt, E., Reid, T., Peterson, S., Baillie, L., McKinney, N., Leighton, T., and J. Hunter-Cevera. 1999. Whole-Genome Sequencing of *Bacillus anthracis*. 1999. DP99, 1st European Conference on Dangerous Pathogens, Winchester, England.

WHOLE-GENOME SEQUENCING OF BACILLUS ANTHRACIS

Timothy READ[1], Scott PETERSON[1], Les BAILLIE[2], Nancy McKINNEY[3], Terry LEIGHTON[3], Jenny C. HUNTER-CEVERA[3] and Eric EISENSTADT[4]

[1]The Institute for Genomic Research, Rockville, Maryland 20850 USA; [2]DERA, Microbial CBD, Porton Down, Salisbury, Wiltshire SP4 0JQ UK; [3]Center for Environmental Biotechnology, Lawrence Berkeley National Laboratory, Berkeley, CA 94720 USA; [4]Office of Naval Research, Arlington, VA 22217-5660 USA

To promote the development of target-based medical countermeasures for *Bacillus anthracis*, The Institute for Genomic Research (TIGR) has initiated an effort to sequence the genome of a virulence-plasmid cured *B. anthracis* Ames strain. The approach, used by TIGR for 10 other microbial genome projects as of July 1999, is to construct a 1.5 - 3.0 kb (small-insert) plasmid library from randomly sheared chromosomal DNA, obtain enough sequence to provide 7 – 8-fold nucleotide coverage of the genome and then assemble the data into contigs (assemblies). To 'close' gaps between assemblies, a combination of PCR techniques, small-insert clone walking and information obtained from end-sequencing large-insert libraries (*e.g.* bacteriophage λ and/or BAC) will be used.

We made a number of libraries in the pUC18 vector, either using *SmaI* blunt ends, or *BstXI* linkers, and, as of 7 June 1999, had generated 15,549 good sequencing reactions from 21,012 attempts (74% efficiency) with an average read length of 568 nt. Based on an estimated genome size of 4.25 Mb, this represented an average coverage of 2.1-fold per base. The average G + C content of this DNA sequence was 36.4%.

Assembly of the *B. anthracis* sequences to date has been in line with the Poisson distribution-based Lander-Waterman model (Genomics 1988 2: 231-239): sequences from the small-insert library have assembled into about 3794 contigs, totaling 4.2 Mb. Thus, our small-insert library is behaving as if it is highly representative of the entire *B. anthracis* genome.

A summary of our preliminary analyses of the *B. anthracis* genomic sequence will be presented.

The *B. anthracis* sequence will be made available through the TIGR microbial database site.

Contact: Dr. Eric Eisenstadt, ONR Code 335, 800 North Quincy Street, Arlington, VA 22217-5660; Tel: (703) 696-4596; Fax: (703) 696-1212

Two most suitable sessions for abstract: **Management of dangerous pathogens**; **Defining dangerous pathogens**

Oral presentation preferred

Molecular Phylogenetic and Ecological Analysis of the *B. anthracis* Clade. T. LEIGHTON*[1], E. EL HELOW[1], M. MOCK[2] and A. FOUET[2]. [1]Division of Biochemistry and Molecular Biology, University of California, Berkeley, California 94720; and [2]Toxines et Pathogénie Bactériennes, Institut Pasteur, 28 rue du Dr Roux, 75724 Paris Cedex 15.

The classical taxonomic relationships between *B. anthracis* and closely related spore forming species such as *B. cereus, B. thuringiensis* and *B. mycoides* are controversial. Spore of bacilli are composed of genetically unique proteinaceous structures. Analogous spore structural proteins only exist in endospore forming bacteria. Several Small Acid Soluble Proteins (SASP) are constituents of the bacterial endospore cytoplasm that protect the spore chromosome from DNA damage. SASP hydrolysis provides amino acids required for protein synthesis during the early stages of spore germination. We have examined the utility of SASPs as molecular chronometers in the phylogenetic reconstruction of spore forming taxa. SASP protein sequences have been used to construct molecular phylogenies of the *Bacillus* lineage using a Maximum Parsimony algorithm. Consistent phylogenetic trees are obtained that are a unique computation solution to the input dataset. These cladograms suggest that *B. anthracis* clade is the most deeply rooted and basal species cluster of the *Bacillus* genus. In cladograms that include anaerobic *Clostridium* lineages, the *B. anthracis* clade is the founding aerobic emergence from the *Clostridium* group. *B. subtilis*, *B. globigii*, *B. globisporus*, *B. stearothermophilus*, and *B. caldolyticus* are terminal taxa in SASP and other cladograms. One SASP gene product, SASP-B (*sspE*), has been identified whose sequence is different in each classically defined *Bacillus* species that we have examined. PCR strategies have been used to obtain full length SASP-B genes from *B. anthracis, B. cereus, B. thuringiensis, B. mycoides,* and other related bacilli. These SASP-B protein sequences are different in each of the classically defined *B. cereus*, *B. thuringiensis, B. mycoides* and *B. anthracis* species. Computational analysis of SASP-B sequence datasets suggests that the *B. anthracis* clade is more phylogenetically complex than suggested by traditional numerical and phenetic classification schemes. Specifically, the classically defined *B. cereus* and *B. thuringiensis* species are not phylogenetically coherent in SASP-B cladograms. SASP-B phylogenetic analysis of environmental isolates from areas of France endemic for anthrax has identified several new *B. anthracis* clade phylotypes that have not been previously recognized as coherent taxonomic units. This research was supported by the Office of Naval Research and the Biology Workbench Project at the National Center for Supercomputing Applications.